**Bacterial endosymbionts in animals**  
Nancy A Moran* and Paul Baumann†

Molecular phylogenetic studies reveal that many endosymbioses between bacteria and invertebrate hosts result from ancient infections followed by strict vertical transmission within host lineages. Endosymbionts display a distinctive constellation of genetic properties including AT-biased base composition, accelerated sequence evolution, and, at least sometimes, small genome size; these features suggest increased genetic drift. Molecular genetic characterization also has revealed adaptive, host-beneficial traits such as amplification of genes underlying nutrient provision.

**Introduction**

In recent years, technological advances allowing molecular phylogenetic characterization have enabled exploration of the world of bacteria that cannot be cultured — a category that includes the majority of life forms. Among the non-cultivable bacteria about which we have discovered most are the endosymbionts that live in animal cells and are transmitted vertically at the time of host reproduction. Although common in many invertebrates, intracellular bacterial associates of animals were little studied until about 10 years ago. A large portion of what was known was compiled in a book by Paul Buchner [1], which remains the central reference for information on the diversity and distribution of endosymbionts. One of the most intriguing aspects of these endosymbioses concerns their evolutionary origins and adaptive modifications.

The bacteria that inhabit animal cells can be divided into several groups. The most distinctive are ‘primary’ symbionts that reside within specialized host cells called bacteriocytes. They have reciprocally beneficial — often reciprocally obligate — relationships with hosts and occur in many terrestrial arthropods as well as some marine invertebrates. ‘Secondary’ symbionts and intracellular pathogens are more sporadically associated with host individuals and vary in tissues occupied. Because effects on hosts are usually unknown, there is no clear demarcation between symbionts and pathogens: a wide range of interactions certainly exist between non-bacteriocyte associates and their hosts (for example see [2†]). Infection of hosts can be strictly maternal, maternal with occasional horizontal transfer, or entirely horizontal, and there is no absolute correspondence between mode of transmission and effects on host fitness.

This review is concentrated on evolutionary aspects of endosymbiosis involving bacteriocyte associates of animals. For these bacteria, there is clear evidence of a coevolved, mutualistic relationship with the host and a distinctive set of genetic traits that result from the association. To date, most studies have focused on insect symbionts, although there are a few evolutionary studies of symbionts in other invertebrate groups. The best-characterized animal endosymbiont is *Buchnera aphidicola*, a bacteriocyte-associated mutualist of aphids, insects that feed on phloem sap of host plants. Many comparative studies of host-beneficial and other loci have been carried out using *Buchnera*, and a full genome has recently been completely sequenced (H Ishikawa, personal communication), with sequencing of another genome in progress. We emphasize molecular studies within the past two years that have applied molecular approaches to reconstruct the evolution of *Buchnera* and other bacteriocyte-associates.

**Molecular phylogenetics and co-speciation**

Although Buchner [1] speculated about the age and origin of endosymbioses, no firm information about this topic was possible for noncultivable symbionts before DNA sequencing became feasible. The first such studies were on *Buchnera*, for which phylogenetic analyses have now revealed matching between phylogenies of symbionts and aphid hosts over a variety of evolutionary time scales. This congruence between host and symbiont phylogenetic trees implies co-speciation and synchronous diversification. Initial studies, for distantly related aphids and their *Buchnera* associates, supported vertical transmission down host lineages from the time of the common ancestor of aphids, which is estimated to be some 150–250 million years ago on the basis of dating from host fossils [3*]. Recently, phylogenetic congruence has been shown for *Buchnera* of closely related aphids that interact ecologically [4]. Analyses of intraspecific polymorphisms in aphid mitochondrial and *Buchnera* markers suggest that horizontal transfer is absent even within a single aphid species [5]. Thus, maternal transmission appears to be the sole mechanism of infection. An implication of these results for bacterial population structure is that no genetic recombination between strains of bacteria from different aphids has occurred.

Phylogenetic congruence with hosts, implying co-speciation, also has been reported for other bacteriocyte-associates, including symbionts of tsetse flies [6*], cockroaches [7], certain marine bivalves [8*], carpenter ants [9*], and the *Wolbachia pipientis* that appear to be mutualistic in nematodes [10†]. All of these studies span ancient divergences among taxa, indicating that, in each case, a single ancient infection was followed by co-speciation across millions of years, with symbiotic bacteria diverging in parallel with their hosts.
Although there is probably some association between maternal transmission as the primary route of infection and congruent of host and symbiont phylogenies spanning evolutionary time periods, either may exist without the other. Some secondary symbionts and intracellular pathogens in insects are transmitted maternally but nonetheless undergo occasional horizontal transmission, through an unknown route, as implied by molecular studies of Wolbachia pipiensis [11,12] and secondary symbionts of aphids [13] and tsetse flies [7]. Conversely, some symbionts undergo co-speciation with hosts in the absence of maternal transmission. This situation appears to characterize Vibriofischeri, which colonize the light organs of squidssquids [14]; these symbionts exist in sea water and reinfest juveniles each generation but show phylogenetic congruence with hosts and adaptation to their own host lineage in experimental transfers [15•]. Symbionts of other marine invertebrates, such as hydrothermal vent tube worms, lack both maternal transmission and phylogenetic congruence with hosts [16].

Molecular phylogenetic analyses also reveal the relationships of endosymbionts to other groups of bacteria [3•]. Several, including the associates of aphids [17••], tsetse flies [6•], psyllids [18•], ants [19] and some bivalves [8•,20], are related to the Flavobacteria in cockroaches [7]. Among the γ-Proteobacteria, endosymbionts of different host groups have evolved as independent lineages from nonsymbiotic bacteria [3•].

The precise phylogenetic relationships of endosymbiotic bacteria are often uncertain, because of limited information in the 16S rRNA gene (the primary sequence used for phylogenetic characterization) and also the elevated base substitution rates and base compositional biases typical of symbionts (21,22• see below). These features generate problems in phylogenetic reconstruction by causing both convergence (sequence identity not due to common ancestry) and ‘long branches’ (taxa so divergent that relationships to other lineages are obscured). Consequently, results from large-scale analyses, such as the Ribosomal Database Project [23], are particularly prone to error in the placement of symbiotic bacteria.

Because host fossils can be used to date symbiont ancestors in cases for which molecular phylogenies support strict co-speciation, sequence divergences within symbiotic clades have been used for calibrating evolutionary rates [3•,6•,8•,24•]. These rates might be used to estimate ages of other bacterial clades, for which direct dating of ancestors is not possible, although rate variation among lineages can make such extensions difficult [25].

**Documenting metabolic contributions of symbionts**

One obvious and long standing question regarding the evolution of symbionts concerns the existence and nature of symbiotic adaptations that benefit the hosts. Most animal symbionts contribute rare nutrients that the host itself cannot make. The biosynthetic contributions of symbionts have been explored in several systems, including Buchnera [26–28], through experiments using symbiotic and artificially cured hosts. In addition, studies of enzymatic activity or end product synthesis by symbionts have been used to document metabolic contributions, such as sulfur oxidation and assimilation of inorganic nitrogen and carbon in symbionts of marine hosts (for example [29••,30,31]). Metabolic studies in intact symbionts have recently been complemented and extended by molecular characterization of genes involved in relevant pathways (for example [17••,32,33,34,35]). The retention of symbiont loci underlying a particular metabolic pathway is strong evidence for a contribution to host metabolism, since bacterial lineages routinely lose genes not used [36].

**Plasmid-borne biosynthetic genes and gene amplification**

The molecular basis of symbiont adaptations for providing nutrients to their hosts is best studied in Buchnera. Plant phloem sap is deficient in essential amino acids, and Buchnera provides several of these to aphid hosts [17••]. Despite its small genome, Buchnera retains genes for biosynthesis of several amino acids that the hosts are unable to produce for themselves. Furthermore, in many Buchnera lineages, genes underlying the rate-limiting step of tryptophan biosynthesis (trpEG) and genes underlying leucine biosynthesis (leuABCD) have been recruited to plasmids [17••]. The leu plasmid is of the IncFII group first isolated from Salmonella and bears, in addition to repA genes characteristic of those plasmids, a single copy of leuABCD, whereas the trpEG plasmid bears tandem repeats of a unit containing trpEG [17••]. In each case, the plasmid location allows amplification relative to chromosomal genes, presumably allowing increased expression and increased benefit to hosts. Several studies address the evolution of these two plasmids and the corresponding pathways [37,38•,39,40•,41]. For both leuABCD and trpEG, these studies support a single recruitment from the ancestral location on the Buchnera chromosome, followed by strictly vertical transmission of the plasmids within Buchnera/aphid lineages. The number of copies of leuABCD and of functional trpEG appear to vary together across Buchnera of different aphid lineages, perhaps reflecting coordinated, adaptive adjustment to nutritional needs of different host species [42•].

Plasmid amplification of amino acid biosynthetic genes presumably evolved as an adaptation benefiting host nutrition. Curiously, this adaptation seems to have degenerated in a few species through the silencing of trpEG repeats as pseudogenes [38•,43]. If host nutritional needs are reduced for some ecological reason, this gene silencing may be adaptive, but it is still not obvious why pseudogene repeats are retained. In the absence of counterbalancing selection, they should be quickly eliminated through homologous
recombination. The explanation may lie in reduced capacity for such recombination in Buchnera. This possibility is supported by the sequencing of the first full Buchnera genome; most recombinase genes are missing, including recA, a locus usually retained even in highly reduced bacterial genomes (H Ishikawa, personal communication). The loss of recombination pathways may represent a host-level adaptation that acts to stabilize the amplification of trpEG for the majority of Buchnera in which the amplification is beneficial.

**Gene sequence evolution and genetic drift in endosymbionts**

The DNA sequence evolution of Buchnera is unusual in several respects when compared with the gene sequence evolution of free living bacteria such as Escherichia coli. First, sequences are very AT-biased (about 28% GC) [17**,44]. This bias is concentrated at the third positions of codons and in spacers but is also present throughout protein-coding sequences where it affects both polypeptide composition [21,24*,45,46*] and gene length [47]. Second, DNA sequences evolve faster in Buchnera than in free-living relatives, especially at sites under natural selection. Several of these features also have been found in other endosymbiotic bacteria [18*,21,22*,48*] and in intracellular pathogenic bacteria (for example [36]).

The most plausible explanation for these patterns is an increased rate of fixation of mildly deleterious mutations within these bacterial populations. This higher rate is expected because symbionts will experience reduced effective population sizes and, consequently, increased levels of genetic drift. Under this view, the enhanced rate of sequence evolution is due to predominantly deleterious change, although some of the substitutions could compensate for effects of these deleterious mutations. This hypothesis is supported by several observations: first, substitutions within rRNA sequences have the effect of destabilizing secondary structure [48*]; second, there is no adaptive codon bias [45,46*]; third, the increase substitution rate in protein-coding genes is concentrated at nucleotide sites that are subject to purifying (conservative) natural selection [21,45,46*]; and fourth, this increased rate is observed at every locus [24*,46*]. This genome-wide pattern is expected if the increase in rate of sequence evolution is related to population size rather than to specific action of selection at particular loci. An alternative explanation for the rate change is relaxation of selection across all or most loci, possibly due to the more constant environment within host cells and/or to the lack of exponential growth phases within the symbiont life cycle. However, even amino acid biosynthetic genes, that are clearly under strong selection in the context of the mutualistic relationship with the host, display elevated rates of evolution.

**Genomic characterization**

In addition to AT-bias and fast sequence evolution, Buchnera shows another distinctive evolutionary trend, that of reductive evolution. On the basis of pulsed field gel electrophoresis it has been shown that Buchnera of the aphid Acyrthosiphon pisum has a small genome at about 650 kb [49**]. Of 130 open reading frames characterized in Buchnera of Schizaphis graminum, all have close homologs in E. coli [17**]. Based on the genome size average gene length of about one kilobase, and the absence of much intergenic spacer between Buchnera genes [17**], this observation suggests that the Buchnera genome consists wholly of a subset of about 600 of the 4500 genes present in an E. coli-like ancestor. The loci published so far correspond to a wide range of basic housekeeping functions in addition to many loci underlying the biosynthesis of essential amino acids, the nutrients required by insect hosts [17**]. The complete genome sequence of Buchnera of A. pisum, to be published this year, will reveal whether Buchnera has any recently acquired genes, but, so far, the evidence suggests much less lability of gene content than in the related enterics or in many other bacteria (for example [50]). Whereas some bacterial lineages show continual turnover in gene content while maintaining a genome of roughly constant size, Buchnera and perhaps other endosymbionts as well as pathogens (e.g. Mycoplasma genitalium) [51] probably evolved largely through genome shrinkage from a free-living ancestor.

Remarkably, a recent study quantifying the DNA content of individual cells demonstrated that each Buchnera contains 50–200 chromosomal copies [52*]. Chromosome copy number appears to vary with the life cycle stage of the host, suggesting chromosome amplification as a way of adaptively varying symbiont contributions to the host nutritional economy.

**Conclusions**

Bacteriocyte associates show long histories with invertebrate hosts, and this life style has resulted in distinctive genetic properties. These result from a combination of deleterious evolution due to increased genetic drift and adaptive evolution in the context of the mutualistic association. The finding that at least some endosymbiont genomes are small has made full genome sequencing feasible, and one Buchnera genome is already complete (H Ishikawa, personal communication). Because Buchnera and some other bacteriocyte-associates are related to E. coli, for which most gene functions are known, full genome sequences promise to present an essentially complete picture of the metabolic potential of symbionts and their capacity for contributing to hosts. If several full genome sequences become available, comparative analysis could be used to explore the processes whereby host-beneficial endosymbiont genes are maintained or lost, through processes of mutation, genetic drift and selection within and between hosts.

In addition to bacteriocyte-associates, recent molecular characterization and in situ hybridizations have revealed
that a diversity of other bacterial groups are present in invertebrates and often maternally transmitted. Their effects on hosts and their patterns of evolution are little known, and will be elucidated during the next few years. So far, much remains unknown, and new studies will reveal unsuspected interactions, for example between symbionts and viruses [53•,54] or between symbionts and hosts (for example [2•]).

Update
Three papers on molecular phylogenetic studies of insect endosymbionts have recently been accepted for publication [55•,56•,57•].

Acknowledgement
We thank Howard Ochman for his comments on a draft of this paper.

References and recommended reading
Papers of particular interest, published within the annual period of review, have been highlighted as:

* of special interest
** of outstanding interest

2. Bandi C, McCall JW, Genchi C, Corona S, Venca L, Sacchi L: Effects of tetracycline on the filarial worms Brugia pahangi and Dirofilaria immitis and their bacterial endosymbionts Wolbachia. Int J Parasitol 1999, 29:357-364. Tetracycline kills the Wolbachia of filarial nematodes and results in embryo degeneration, suggesting that this association may be mutualistic. This contrasts with the Wolbachia in insects, which are reproductive parasites.
3. Moran NA, Telang A: Bacteriocyte-associated symbionts of insects – a variety of insect groups harbor ancient prokaryotic endosymbionts. Bioscience 1998, 48:295-304. This article summarizes molecular phylogenetic information on different bacterial symbionts of insects and reports estimates of how long symbiont–host associations have been in existence.
4. Clark MA, Moran NA, Baumann P, Wernegreen JJ: Consecration between bacterial endosymbionts (Buchnera) and a recent radiation of aphids (Uroleucon) and pitfalls of testing for phylogenetic congruence. Evolution 2000, 54:517-525.
6. Chen XA, Li S, Aksoy S: Concordant evolution of a symbiont with its host insect species: molecular phylogeny of genus Glossina and its bacteriome-associated endosymbiont, Wigglesworthia gossinii. J Mol Evol 1999, 48:49-58. The authors show that parallel evolution of the tsetse primary endosymbiont and the host is consistent with a single infection of the host followed by strictly vertical transmission of the endosymbiont. The secondary endosymbionts, on the other hand, are vertically identical indicating either horizontal evolution or multiple independent infections with the same organisms. This pattern parallels that for aphids and their primary and secondary symbionts.
15. Nishiguchi MK, Ruby EG, McFall-Ngai MJ: Competitive dominance among strains of luminous bacteria provides an unusual form of evidence for parallel evolution in Sepiolid squid-vibrio symbioses. Appl Environ Microbiol 1998, 64:3209-3213. This study suggests that specialization to particular host lineages and vertical transfer down host lineages can occur even when infection by symbionts is acquired horizontally each generation.
22. Peek AS, Vrijenhoek RC, Gauth BS: Accelerated evolutionary rate in sulfur-oxidizing endosymbiotic bacteria associated with the mode of symbiont transmission. Mol Biol Evol 1999, 15:1514-1523. Patterns of sequence evolution suggesting deleterious evolution, previously pointed out for insect endosymbionts [21], also extend to vertically transmitted bivalve endosymbionts but not to symbionts that infect later during development of the host. The results suggest that these genetic features may have broad occurrence in animal symbioses.


This is the largest comparative study of endosymbionts to date, based on number of loci examined. Sequences for 24 loci are obtained and analyzed for four Buchnera taxa, comprising two species pairs each diverging an estimated 50 million years ago. For all chromosomal loci, synonymous divergences for a given taxon pair are similar across loci, suggesting homogenous mutation rates; however, loci on the trpEG plasmid experience twice as much mutation. The synonymous substitution rate, which approximates the mutation rate, is estimated at about 0.2 substitutions per site per billion generations, about twice the value estimated based on synonymous divergences for enteric bacteria. Reconstruction of DNA sequence evolution at protein-coding genes indicates that most of the AT-bias evolved in a common ancestor of extant Buchnera.


33. Baumann L, Baumann P, Thao ML: Carbon release from purified recombinant enzymes of the leucine plasmid carried by the endosymbiont (Buchnera aphidicola) from aphids of the family Aphididae; in this paper, a new plasmid type is described from Diuraphis noxia. Appl Environ Microbiol 1999, 65:17-25.


42. Thao ML, Baumann L, Baumann P, Moran NA: Endosymbionts (Buchnera) from the aphids Schizaphis graminum and Diuraphis noxia have different copy numbers of the plasmid containing the leucine biosynthetic genes. Curr Microbiol 1998, 36:238-240.


Chromosomal amplification up to 300-fold is inferred by estimating amount of DNA per cell and dividing by the chromosome size.


Some of the GroEL protein produced by Buchnera in aphids is transferred to the haemolymph where it binds to certain viruses allowing them to persist in the host. Thus, the endosymbionts play a role in the capacity of the aphid to vector plant viral diseases.


Molecular phylogenetic studies have shown that, in the Buchnera of the aphid genus Diuraphis, plasmid-borne trpEG pseudogenes have evolved multiple times, and they persist in Buchnera of D. noxia collected from geographically diverse populations. These pseudogenes appear to represent a reduction of symbiotic contributions based on the observation of unusually low levels of circulating tryptophan in these hosts compared to insects in which Buchnera retain only functional trpEG copies.


Genome size is remarkable conserved within Buchnera suggesting that there is little or no turnover in gene content of Buchnera genomes.


This is the most extensive study to date for the bacteriocyte-associates of psyllids, a group of plant sap-feeding insects. Full 16S-23S rDNA sequences for symbionts of 28 psyllid species indicate a well defined clade of bacteria with extreme AT-bias and fast sequence evolution. Phylogenetic trees based on host nuclear gene sequences show the same branching pattern as those based on bacterial genes.